Molecular Basis for Allelic Polymorphism of the Maize *Globulin-1* Gene

Faith C. Belanger[1] and Alan L. Kriz

Department of Agronomy, University of Illinois, Urbana, Illinois 61801

Manuscript received February 1, 1991
Accepted for publication August 2, 1991

ABSTRACT

An abundant protein in maize (*Zea mays* L.) embryos is a storage globulin encoded by the polymorphic *Glb1* gene. Several *Glb1* protein size alleles and a null allele have been described. Here we report the isolation and nucleotide sequence analysis of genomic clones corresponding to two *Glb1* size alleles (*Glb1-L* and *Glb1-S*) and to the *Glb1–0* null allele. The *Glb1-L* and *Glb1–0* alleles differ from *Glb1-S* by the presence of small nucleotide insertions which are imperfect or perfect duplications, respectively, of adjacent sequences. In the case of *Glb1-L*, the insertion is in-frame and results in a protein larger than that encoded by *Glb1-S*, whereas in *Glb1–0* the insertion causes a translational frameshift which introduces a premature termination codon. Although steady-state levels of *Glb1–0* transcripts are extremely low in *Glb1–0/0* embryos, nuclear transcription assays indicate that the *Glb1–0* gene is transcribed at a level comparable to that of *Glb1-L*. This suggests that the low amounts of *Glb1–0* transcripts in the cytoplasm may be due to mRNA instability.

GLOBULINS are the major storage proteins in maize embryos, accounting for 10–20% of the embryo protein (KRIZ 1989). The major globulin component, GLB1, is one of the most abundant proteins in normal mature embryos. GLB1 is encoded by the single gene *Globulin-1* (*Glb1*) for which several size alleles and a CRM^- (cross-reacting material) null have been described (SCHWARTZ 1979; OSTERMAN 1988). The three most common *Glb1* alleles have been designated *L*, *I*, and *S* for Large, Intermediate, and Small proteins, respectively. Several characteristics of *Glb1* make it an interesting system for study: (1) GLB1 is an abundant protein encoded by a single gene (SCHWARTZ 1979), (2) expression of the *Glb1* gene is seed specific (BELANGER and KRIZ 1989) and (3) GLB1 protein is not essential for seedling growth since homozygosity for the *Glb1–0* null allele has no effect on embryo development, maturation, or subsequent germination (SCHWARTZ 1979).

We previously reported the characterization of a cDNA clone for the *Glb1-S* allele (BELANGER and KRIZ 1989). Here we report the isolation and characterization of genomic clones for the *Glb1-S*, *-L* and *-0* alleles.[2] Analysis of these clones has revealed the nature of allelic polymorphisms in *Glb1*. Nucleotide sequence comparisons indicate that the *Glb1-L* and *Glb1–0* alleles are more closely related to each other than is either allele to *Glb1-S*, and it appears that the *S* allele is the progenitor from which the other two alleles are derived. Both the *Glb1-L* and *0* alleles differ from the *S* allele by small insertions within their respective protein coding sequences. In *Glb1-L* the insertion is in frame and results in a larger protein, whereas in the *Glb1–0* allele the insertion causes a frameshift mutation in the amino-terminal region of the protein which is followed shortly by a premature termination codon. Premature termination of translation of *Glb1–0* mRNA apparently results in transcript instability: from nuclear run-on experiments the level of transcription from *Glb1–0* is similar to that of *Glb1-L* although the steady state level of the *Glb1–0* transcript is barely detectable by northern blot analysis (BELANGER and KRIZ 1989).

MATERIALS AND METHODS

Materials: Embryos homozygous for the *Glb1-L* and *Glb1-S* alleles were obtained from field-grown plants of the maize (*Zea mays* L.) inbred lines W64A and Va26, respectively, as previously described (KRIZ 1989). The *Glb1–0* allele was originally identified in a Black Beauty popcorn line (SCHWARTZ 1979) and is maintained in a homozygous state by a combination of plant outcrossing and selfing. Embryos were harvested as previously described and frozen in liquid N_2. LambdaZAP vector arms, Gigapack packaging extracts, and exonuclease III/mung bean nuclease deletion kits were from Stratagene (La Jolla, California). α-^{32}P-Labeled dATP (3000 Ci/mmol) was obtained from New England Nuclear (Boston, Massachusetts). NA45 paper was obtained from Schleicher and Schuell (Keene, New Hampshire). Gel-X tubes were obtained from Genex (Gaithersburg, Maryland).

Protein extraction and immunoblot analysis: Preparation of protein extracts from mature maize embryos, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and immunoblot analysis were performed as previously described (BELANGER and KRIZ 1989; PUCKETT and KRIZ 1991).

Nucleic acid isolation: Preparation of total DNA from

[1] Current address: Department of Crop Science, Cook College, Rutgers University, New Brunswick, New Jersey 08903.

[2] The EMBL accession numbers for the *Glb1-S*, *-L*, and *-0* sequences reported here are X59084, X59083, and X59085, respectively.

unfertilized ears was as described by DELLAPORTA, WOOD and HICKS (1983) followed by further purification on CsCl gradients. Total RNA was isolated from frozen tissue by using the guanidine-HCl method described by COX (1968). Polyadenylated RNA was fractionated from total RNA by oligo(dT)-cellulose chromatography (AVIV and LEDER, 1972). For use in hybridizations, cDNA fragments from appropriate clones were isolated by *Eco*RI digestion, separation on a 1% agarose gel, and binding of the fragment to NA-45 paper. DNA fragments were labeled with [α-^{32}P] dATP by using a commercial random priming kit (BRL or Stratagene).

Isolation and characterization of genomic clones: To obtain a genomic clone corresponding to *Glb1-L*, a genomic library prepared from W64A nuclear DNA in the lambda vector Charon 32 (KRIZ, BOSTON and LARKINS 1987) was screened by using the *Glb1-S* cDNA clone as a radiolabeled probe essentially as described by HUYNH, YOUNG and DAVIS (1985). Growth of recombinant phage in liquid culture and lambda DNA preparation were performed as previously described (KRIZ, BOSTON and LARKINS 1987). A single clone, designated λgGlb1-L, of 17 kb was found to contain a 3.5-kb *Eco*RI fragment which hybridized with the pcGlb1S probe. This fragment was cloned into the plasmid vector pBluescript SK and given the designation pgGlb1-L.

Genomic clones for *Glb1-S* and *Glb1–0* were obtained by preparing size-selected libraries in the vector LambdaZAP. Southern blots of *Eco*RI digested DNA isolated from unfertilized ears of plants homozygous for each of the three alleles indicated that in all cases a 3.5-kb *Eco*RI fragment hybridized to the *Glb1* probe (see Figure 1C). The DNA from this region of an agarose gel was purified using NA45 paper or Gel-X tubes as suggested by the manufacturers. This DNA was ligated to *Eco*RI-digested LambdaZAP arms and packaged by using the Gigapack system. The resulting libraries were screened by using the radiolabeled *Glb1-S* cDNA clone as probe. The genomic clones were excised from LambdaZAP as recombinant pBluescript SK(−) plasmids according to the manufacturer's protocols. For nucleotide sequence analysis, the genomic clones were subcloned into M13mp18 and mp19 (YANISCH-PERRON, VIEIRA and MESSING 1985) to obtain inserts in opposite orientations. Overlapping unidirectional deletions corresponding to either strand were prepared from the appropriate M13 clone RF by using a commercial exonuclease III/mung bean nuclease deletion kit (Stratagene). Dideoxynucleotide sequencing (SANGER, NICKLEN and COULSEN 1977) of single-stranded templates with T7 DNA polymerase was performed by using commercial sequencing kits (United States Biochemical Corp., Cleveland, Ohio; Pharmacia LKB Biotechnology Inc., Piscataway, New Jersey). The deoxyguanine triphosphate (dGTP) analog 1-deaza dGTP was used to resolve GC compressions. Analysis of DNA sequences was performed on an IBM PC AT with either IBI Pustell Sequence Analysis software (International Biotechnologies Inc., New Haven, Connecticut) or DNAStar programs (DNAStar, Inc., Madison, Wisconsin).

Mapping of 5′ ends of transcripts: Primer extension analysis (KINGSTON 1989) was used to determine the 5′ end of transcripts corresponding to each *Glb1* allele. A 27-base oligonucleotide homologous to the transcribed region from position 133 to 159 in the pcGlb1S cDNA clone was end-labeled with [γ-^{32}P]ATP by using a commercial kit (BRL). The labeled primer was annealed to 1 μg of poly(A$^+$) RNA isolated from 24 days after pollination (DAP) embryos homozygous for either the *S* or *L* alleles. To compensate for the low steady state level of *Glb1* transcripts in *Glb1–0/0* embryos (BELANGER and KRIZ 1989), 16 μg of poly(A$^+$) RNA was used in this case. The primers were extended using reverse transcriptase, the RNA was digested with RNaseA, and the resulting products were electrophoresed on a 6% sequencing gel adjacent to sequencing reactions of the relevant genomic clones primed with the same oligonucleotide.

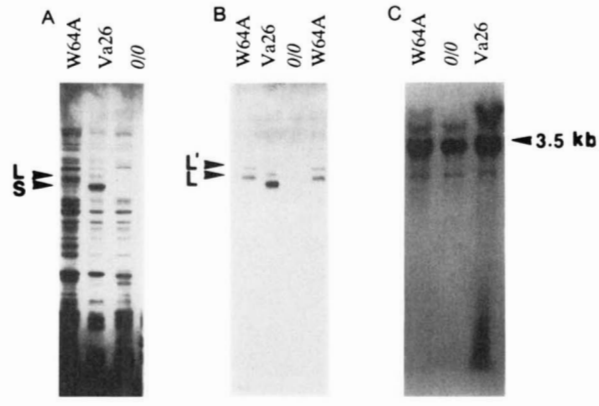


FIGURE 1.—Comparison of *Glb1-L*, *Glb1-S* and *Glb1–0* alleles. A, Coomassie-stained SDS-polyacrylamide gel of proteins extracted from embryos of the inbred lines W64A (*Glb1*-L/L), Va26 (*Glb1-S/S*), and embryos homozygous for the *Glb1–0* allele. B, Immunoblot of identical samples probed with GLB1-specific antiserum. C, Southern blot of DNA extracted from the same maize lines probed with the *Glb1-S* cDNA clone.

Isolation of nuclei and run-on transcription assays: Nuclei were isolated from 4–5 g of frozen embryos as described by BEACH *et al.* (1985). The final nuclear suspension contained a significant amount of starch which was considered excluded volume. The volume of nuclei was determined by centrifuging a 50-μl aliquot of the suspension and measuring the supernatant volume (KODRZYCKI, BOSTON and LARKINS, 1989).

The transcription reaction was essentially as described by BEACH *et al.* (1985). The nuclei were added to concentrated stock solutions to yield a final reaction composition of 0.35 mM ATP, 0.35 mM CTP, 0.35 mM UTP, 4.3 μM GTP, 50 mM $(NH_4)_2SO_4$ and 500 μCi of [α-^{32}P]GTP. Nuclear RNA extraction and hybridization to cloned cDNA fragments were as described by KODRZYCKI, BOSTON and LARKINS (1989). The labeled RNA was hybridized to nitrocellulose filter discs to which were previously bound 1.5 μg single-stranded M13 cDNA clones corresponding to either the coding or noncoding strands of *Glb1*, *Glb2*, which encodes an M_r 45,000 embryo globulin (WALLACE and KRIZ 1991), and *L3*, which encodes the major lipid body protein of maize embryos (VANCE and HUANG 1987). The hybridization for each clone was done in triplicate. After washing (KODRZYCKI, BOSTON and LARKINS 1989), the individual discs were counted by liquid scintillation spectroscopy. The average counts for the noncoding strand were subtracted from the average counts for the relevant coding strand.

RESULTS

Allelic polymorphism of *Glb1*: The nature of allelic polymorphism with respect to size of GLB1 protein was originally described by SCHWARTZ (1979). This polymorphism is apparent in SDS-PAGE and immunoblot analysis of protein extracts from embryos homozygous for the Large (inbred line W64A), Small (inbred line Va26), and null *Glb1* alleles (Figure 1, A

and B). The abundant GLB1 protein is readily detected in a Coomassie-stained gel of total protein extracts from embryos homozygous for either the *Glb1-S* or *-L* allele (Figure 1A). No corresponding protein is present in embryos homozygous for the *Glb1–0* null allele. An immunoblot of identical samples is shown in Figure 1B. The protein processing intermediates GLB1′-L and GLB1′-S are readily detectable in addition to the mature GLB1 protein. Previous studies have indicated that at least three processing steps are involved in the formation of the mature protein from the primary translation product (KRIZ and SCHWARTZ 1986).

Isolation and characterization of genomic clones for *Glb1-S*, *-L*, and *-0* alleles: Isolation of genomic clones corresponding to *Glb1-S* and *Glb1–0* was facilitated by the presence of a single *Glb1*-specific *Eco*RI restriction fragment in the genome. Figure 1C depicts a Southern blot of *Eco*RI digested DNA isolated from plants homozygous for each of the three alleles probed with the *Glb1-S* cDNA insert. All three genotypes contain a single major band at about 3.5 kilobases (kb). This fragment was cloned from DNA of plants homozygous for each genotype and the resultant clones were given the designations pgGlb1-S, pgGlb1-L and pgGlb1–0.

A comparison of the nucleotide sequences of these three clones is shown in Figure 2. Gaps have been inserted to facilitate sequence alignments. In each case the cloned fragment is comprised of about 360 base pairs (bp) 5′ to the coding region, about 2400 bp of coding region (exons plus introns) and about 780 bp of 3′ flanking sequence. The present sequence analysis revealed one error in the published cDNA sequence. The T at position 550 in the *S* allele in Figure 2 was reported as a G in the cDNA sequence (BELANGER and KRIZ 1989). This results in a change from a glycine codon to a cysteine codon.

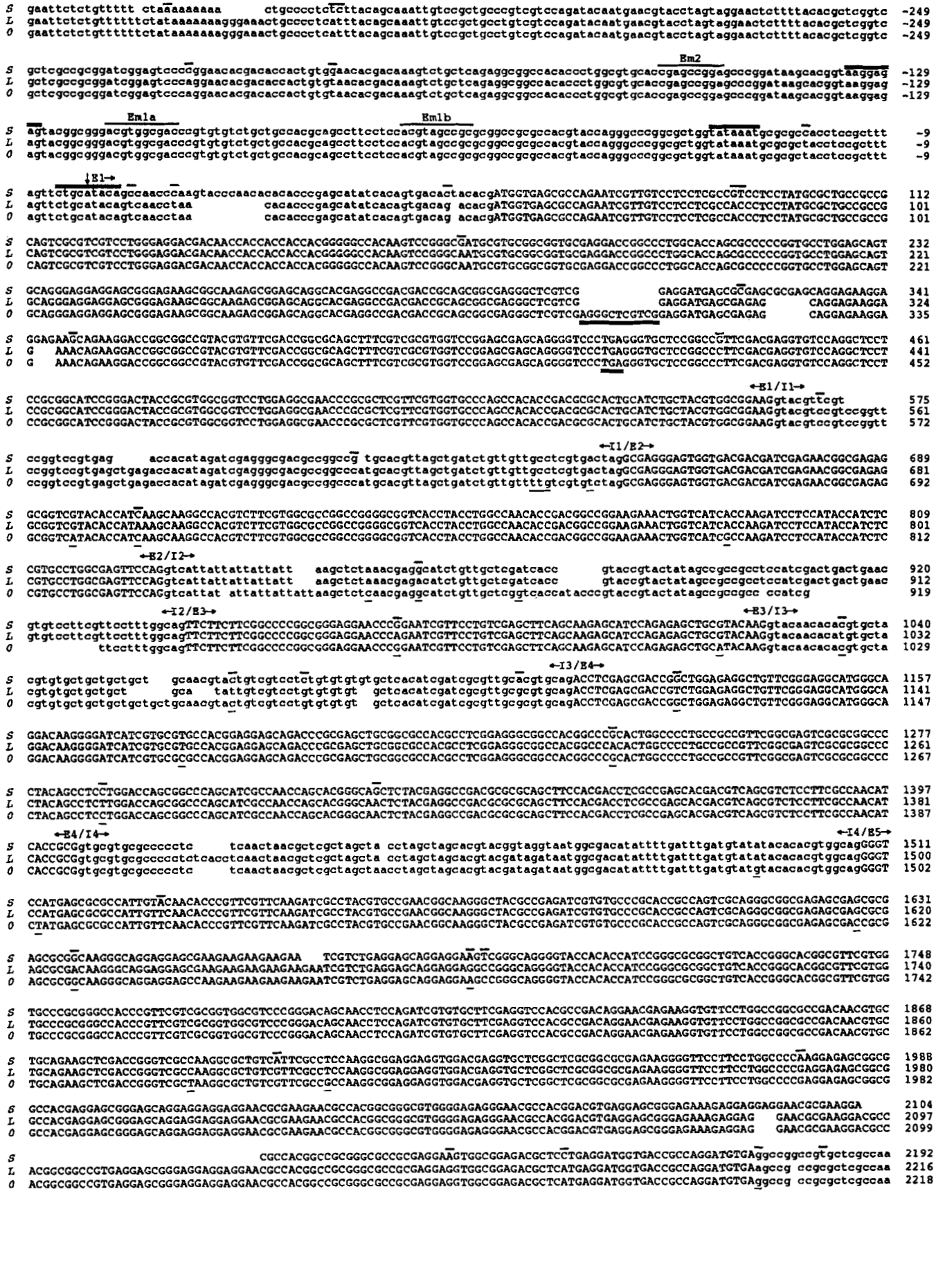
Comparison of the *Glb1-S* cDNA sequence (BELANGER and KRIZ 1989) with the sequences of the genomic clones indicates that each *Glb1* allele contains five exons and four introns, as summarized in Figure 3. The basic organization of all three genes is identical. The 5′ and 3′ exon splice sites are in agreement with the GT/AG consensus observed for all known introns (BREATHNACH and CHAMBON 1981). The small size of the introns, which range from 81 to 114 bp, is characteristic of introns in plant genes (WALBOT and MESSING 1988).

The basis for the size polymorphism observed by SDS-PAGE for *Glb1-L* and *Glb1-S* was determined by comparison of the nucleotide sequences of the two alleles. The difference between *Glb1-S* and *Glb1-L* is largely due to a 36 bp insertion in the last exon beginning at position 2094 in the *L* allele. This insertion is an imperfect duplication of the sequence immediately preceding this region (positions 2049 to 2093). Possible origins for such a duplication are discussed below. Previous peptide mapping studies indicated that the protein size polymorphism could be attributed to a difference at one terminus of the protein (KRIZ and SCHWARTZ 1986). The present sequence analysis demonstrates that the size difference is at the carboxyl terminus of the protein, where GLB1-L contains an additional 12 amino acids relative to GLB1-S (Figure 4).

No GLB1 protein is detectable in embryos homozygous for the *Glb1–0* allele (SCHWARTZ 1979) (also see Figure 1B). Sequence analysis of pgGlb1–0 revealed the presence of an 11-bp insertion, relative to the other two alleles, in the first exon beginning at nucleotide position 299. An 11-bp insertion in the protein coding region necessarily results in a translational frameshift. This frameshift results in a premature termination codon beginning at nucleotide position 413 (Figure 4). The predicted size of the polypeptide which could be translated from *Glb1–0* transcripts is 13.8 kD. No immunoreactive polypeptides of this size have been detected in *Glb1–0/0* embryo extracts using antibodies raised against GLB1 (BELANGER and KRIZ 1989) or GLB1′ (KRIZ and SCHWARTZ 1986). The 11-bp insertion in the *Glb1–0* gene is a perfect duplication of the immediate 5′ sequence.

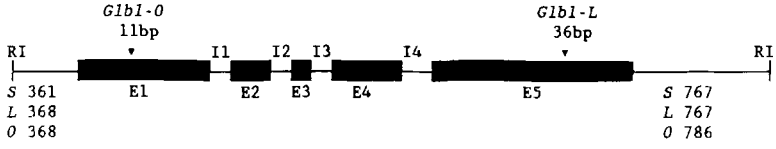
Primer extension analysis was performed to determine the 5′ end of transcripts corresponding to each allele. Previous Northern blot analysis indicated that the steady-state level of *Glb1* transcripts in *Glb1–0/0* embryos was extremely low relative to that of embryos homozygous for either *Glb1-L* or *Glb1-S* (BELANGER and KRIZ 1989). We therefore used a 16-fold excess poly(A^+) RNA from *Glb1–0/0* embryos to obtain a signal comparable to that obtained from the *Glb1-L* and *-S* alleles. By comparing the size of the major primer extension products with sequencing ladders obtained by using the same primer (Figure 5), the 5′ ends of *Glb1* transcripts were determined to be at the same adenine in all three alleles (indicated by an arrow in Figure 2). Minor primer extension products three bases 5′ to the major product were also seen in Va26 and W64A. Because of the higher background it was not possible to determine if a minor product was present in the *Glb1–0/0* embryos. The observation that the major primer extension product is the same in all three alleles confirms that the hybridization signal seen in northern blots of RNA from *Glb1–0/0* embryos is actually due to *Glb1* transcripts.

Transcription of *Glb1* genes: Since the steady-state level of *Glb1* transcripts was low in embryos homozygous for the *Glb1–0* allele, it was of interest to determine how the level of transcription from *Glb1–0* compared with that of a functional *Glb1* allele. We there-


Em2
Em1a
Em1b
E1
E1/I1
I1/E2
E2/I2
I2/E3
E3/I3
I3/E4
E4/I4
I4/E5

```
S  aacgagca    ggaagcaacgagagggtggcgcgcgaccgacgtgcgtacgtagcatgagcctgagtggagacgttggacgtgtatgtatatacctctctgcgtgttaactatgtacgt  2308
L  aacgagcaagcaggaagcaacgagaggatggcgcgcgaccgacgtgcgtacgtagcatgagcctgagtggagacgttggacgtgtatgtatatacctctctgcgtgttaactatgtacgt  2336
0  aacgagcaagcaggaagcaacgagaggatggcgcgcgaccgacgtgcgtacgtagcatgagcctgagtggagacgttggacgtgtatgtatatacctctctgcgtgttaactatgtacgt  2338

                                                                                                                  ←E5↓
S  aagcggcaggcagtgcaataagtgtggctctgtagtatgtacgtgcgggtacgatgctgtaagctactgaggcaagtccataaataaataatgacacgtgcgtgttctataatctcttcg  2428
L  aagcggcaggcagtgcaataagtgtggctctgtagtatgtacgtgcgggtacgatgctgtaagctactgaggcaagtccataaataaataatgacacgtgcgtgttctataatctcttcg  2456
0  aagcggcaggcagtgcaataagtgtggctctgtagtatgtacgtgcgggtacgatgctgtaagctactgaggcaagtccataaataaataatgacacgtgcgtgttctataatctcttcg  2458

S  cttcttcatttgtccccttgcggagtttggcatccattgatgccgttacgctgagaacagacacagcagacgaaccaaaagtgagttcttgtatgaaactatgacccttcatcgctaggc  2548
L  cttcttcatttgtccccttgcggagtttggcatccattgatgccgttacgctgagaacagacacagcagacgaaccaaaagtgagttcttgtatgaaactatgacccttcatcgctaggc  2576
0  cttcttcatttgtccccttgcggagtttggcatccattgatgccgttacgctgagaacagacacagcagacgaaccaaaagtgagttcttgtatgaaactatgacccttcatcgctaggc  2578

S  tcaaacagcaccccgtacgaacacagcaaattagtcatctaactattagcccctacatgtttcagacgatacataaata                                          2627
L  tcaaacagcaccccgtacgaacacagcaaattagtcatctaactattagcccctacatgtttcagacgatacataaata                                          2655
0  tcaaacagcaccccgtacgaacacagcaaattagtcatctaactattagcccctacatgttccagacgatacataagttaataggagtgtgcatccctaccaatgaactattatagccct  2698

S  tagcccatccttagcaattagctattggccctgcccatcccaagcaatgatctcgaagtatttttaatatatagtatttttaatatgtagcttttaaaattagaagataatttttgagaca  2747
L  tagcccatccttagcaattagctattggccctgcccatcccaagcaatgatctcgaagtatttttaatatatagtatttttaatatgtagcttttaaaattagaagataatttttgagaca  2775
0  tagcccatccttagcaatgagctattggccctgcccatc                    agtatttttaatatatagtatttttaatatatctttaaaattagtagataatttttgaaaca  2801

S  aaaatctccaagtatttttttgggtatttttttactgcctccgttttttctttatttctcgtcacctagtttaattttgtgctaatcggctataaacgaaacagagagaaaagttactctaa  2867
L  aaaatctccaagtatttttttgggtatttttttactgcctccgttttttctttatttctcgtcacctagtttaattttgtgctaatcggctataaacgaaacagagagaaaagttactctaa  2895
0  aaa tctccaagtatttttttctggtatttttttactgcctccgttttttatttatttctcgtcacctagtttaattttgtgctaatcggctaaaaacgaaacagggagaaaagttactctaa  2920

S  aagcaactccaacagattagatataaatcttatatcctgcctagagctgttaaaaagatagacaactttagtggattagtgtatgcaacaaactctccaaatttaagtatcccaactacc  2987
L  aagcaactccaacagattagatataaatcttatatcctgcctagagctgttaaaaagatagacaactttagtggattagtgtatgcaacaaactctccaaatttaagtatcccaactacc  3015
0  aagcaactccaacagattagatataaatc      ctgcctagagctgttaaaaagatagacaagtttagtggattagtgtatgcaacaaactctccaaatttaagtatcccaaaaacc  3033

S  caacgcatatcgttccctttcattggcgcacgaactttcacctgctatagccgacgtacatgttcgtttttttt   gggcggcgcttactttcttccccgttcgttctcagcatcgca  3104
L  caacgcatatcgttccctttcattggcgcacgaactttcacctgctatagccgacgtacatgttcgtttttttt   gggcggcgcttgctttcttccccgttcgttctcagcatcaca  3132
0  caacgcatatcgttccctttcattggcgcatgaactttcatctgctatagccgacgtacatgttcgttttttttttttgcggcacttgctttcttcaccgttcgttctcagcatcgca  3153

S  actcaatttgttatggcggagaagcccttgtatcccaggtagtaatgcacagatatgcattattattattcataaaagaattc                                  3187
L  actcaatttgttatggcggagaagcccttgtatcccaggtagtaatgcacagatatgcatcattattattcataaaagaattc                                  3215
0  actcaatttgttatggctgagaagcccgtgtatcccaggtagtaatgcacagatatgcattattattattcataaaagaattc                                  3236
```

FIGURE 2.——Nucleotide sequence comparison of the *Glb1-S*, *Glb1-L* and *Glb1-0* genes. Gaps have been inserted to facilitate alignment. Numbering of each sequence is relative to the transcription initiation site as determined by primer extension analysis (Figure 5). The sequence of the *L* allele is used as the standard for comparisons. Single base differences in the *S* sequence relative to the *L* sequence are overlined, and differences in the *0* sequence relative to the *L* sequence are underlined. Noncoding flanking sequences, 5′ and 3′ untranslated regions, and introns are in lower case, and protein coding regions are in upper case. The transcription initiation site (position 1) and polyadenylation site (position 2420 in pgGlb1S), as determined from analysis of the pcGlb1S cDNA clone (BELANGER and KRIZ 1989), are indicated by arrows. The consensus sequence surrounding the transcription initiation site (position −4 to 5) is indicated by a double overline. The 5′ and 3′ boundaries of each exon (E) and intron (I) are indicated above the *Glb1-S* sequence. The 11 bp duplication in *Glb1-0* (position 299–309) and the resulting in-frame premature termination codon (position 413–415) are both double underlined. The sequences similar to the Em1a, Em1b, and Em2 putative ABA-responsive elements (MARCOTTE, RUSSELL and QUATRANO 1989) are as indicated. Other putative regulatory sequences, indicated by double overlines, are the AGGA element (potentially involved in seed-specificity of expression) at position −134 to −127 and the TATA element at position −33 to −27.



Exon:	1	2	3	4	5
S	565	176	81	291	913
L	545	176	81	291	952
0	556	176	81	291	952

Intron:	1	2	3	4
S	87	114	90	103
L	99	114	82	108
0	99	100	91	104

FIGURE 3.—Structural features of *Glb1* alleles. The organization of the *Glb1-S* gene with respect to exons (dark boxes), 5′ flanking sequence, introns, and 3′ flanking sequence (lines) is diagrammed as the reference allele. Relative positions of insertions in the *Glb1-0* and *Glb1-L* genes are indicated. Lengths of flanking sequences, exons, and introns for each allele are presented in base pairs. RI, *Eco*RI; E, exon; I, intron.

fore performed *in vitro* run-on transcription assays. Nuclei were isolated from 24 DAP *Glb1-L/L* embryos and *Glb1-0/0* embryos. The nuclei were added to an *in vitro* run-on transcription reaction and the resulting RNA was used to probe nitrocellulose filter dots to which were bound single-stranded DNA from cDNA clones corresponding to *Glb1* (BELANGER and KRIZ 1989), the M_r 45,000 maize embryo globulin GLB2 (WALLACE and KRIZ, 1991), and the lipid body protein L3 (VANCE and HUANG, 1987). In order to compare the level of *Glb1* transcription in the two genotypes, the counts obtained for *Glb1* were normalized to the counts obtained for *Glb2* in each genotype. The transcription level of *Glb2* was chosen as a standard since the steady state level of *Glb2* transcripts in both the *Glb1-L/L* and *Glb1-0/0* genotypes appeared to be similar (Figure 6A). After normalization to the transcription level of *Glb2*, the *Glb1* transcription level in *Glb1-0/0* embryos was 77% of that determined for *Glb1-L/L* embryos. This high transcription level indi-

```
GLB1S   MVSARIVVLLAVLLCAAAAVASSWEDDNHHHHGGHKSGRCVRRCEDRPWHQRPRCLEQCREEEREKRQERSRHEADDRSGEGSSEDEREREQEKEEKQKDR
GLB1L   MVSARIVVLLATLLCAAAAVASSWEDDNHHHHGGHKSGQCVRRCEDRPWHQRPRCLEQCREEEREKRQERSRHEADDRSGEGSSEDERE--QEKE-KQKDR
GLB10   MVSARIVVLLATLLCAAAAVASSWEDDNHHHHGGHKSGQCVRRCEDRPWHQRPRCLEQCREEEREKRQERSRHEADDRSGEGSSrarrrmsesrrrrnrrt

GLB1S   RPYVFDRRSFRRVVRSEQGSLRVLRPFDEVSRLLRGIRDYRVAVLEANPRSFVVPSHTDAHCICYVAEGEGVVTTIENGERRSYTIKQGHVFVAPAGAVTY
GLB1L   RPYVFDRRSFRRVVRSEQGSLRVLRPFDEVSRLLRGIRDYRVAVLEANPRSFVVPSHTDAHCICYVAEGEGVVTTIENGERRSYTIKQGHVFVAPAGAVTY
GLB10   ggrtcstgaafvawsgasrgp♦

GLB1S   LANTDGRKKLVITKILHTISVPGEFQFFFGPGGRNPESFLSSFSKSIQRAAYKTSSDRLERLFGRHGQDKGIIVRATEEQTRELRRHASEGGHGPHWPLPP
GLB1L   LANTDGRKKLVITKILHTISVPGEFQFFFGPGGRNPESFLSSFSKSIQRAAYKTSSDRLERLFGRHGQDKGIIVRATEEQTRELRRHASEGGHGPHWPLPP

GLB1S   FGESRGPYSLLDQRPSIANQHGQLYEADARSFHDLAEHDVSVSFANITAGSMSAPLYNTRSFKIAYVPNGKGYAEIVCPHRQSQGGESERERGKGRRSEEE
GLB1L   FGESRGPYSLLDQRPSIANQHGQLYEADARSFHDLAEHDVSVSFANITAGSMSAPLFNTRSFKIAYVPNGKGYAEIVCPHRQSQGGESERERDKGRRSEEE

GLB1S   EE-SSEEQEEVGQGYHTIRARLSPGTAFVVPAGHPFVAVASRDSNLQIVCFEVHADRNEKVFLAGADNVLQKLDRVAKALSFASKAEEVDEVLGSRREKGF
GLB1L   EEESSEEQEEAGQGYHTIRARLSPGTAFVVPAGHPFVAVASRDSNLQIVCFEVHADRNEKVFLAGADNVLQKLDRVAKALSFASKAEEVDEVLGSRREKGF

GLB1S   LPGPKESGGHEEREQEEEEREERHGGRGERERHGREEREKEEEEREG-----------RHGRGRREEVAETLLRMVTARM
GLB1L   LPGPEESGGHEEREQEEEEREERHGGRGERERHGREEREKEE-EREGRHGGREEREEEERHGRGRREEVAETLMRMVTARM
```

FIGURE 4.—Alignment of deduced amino acid sequences for the GLB1-S, GLB1-L and GLB1-0 proteins. Gaps have been inserted to facilitate alignment. Asterisks indicate amino acid changes between the GLB1-S and GLB1-L sequences. The potential signal sequence cleavage site is indicated by an arrow, and the position of *Mep*-catalyzed cleavage, which produces the mature form of the protein (BELANGER and KRIZ 1989), is indicated by an arrowhead (▼). The altered sequence of the GLB1-0 protein, as conditioned by a frameshift mutation, is indicated in lower case. The position of the premature termination codon in the GLB1-0 sequence is indicated by a diamond (♦).

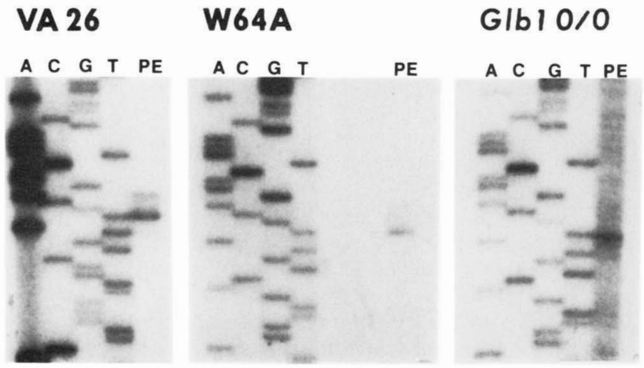


FIGURE 5.—Determination of 5′ ends of transcripts encoded by *Glb1* alleles. Primer extension analysis of poly(A$^+$) RNA isolated from embryos of the inbred lines Va26 (*Glb1-S/S*), W64A (*Glb1-L/L*) and from embryos homozygous for *Glb1-0/0* was performed by using an end-labelled oligonucleotide corresponding to positions 133–159 in *Glb1-S*. The primer extension products were run adjacent to sequencing ladders of appropriate M13 clones corresponding to each allele which were primed with the same oligonucleotide.

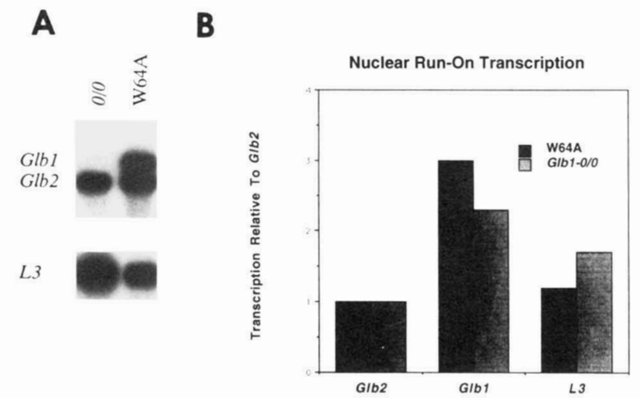


FIGURE 6.—A, Northern blot analysis of 1 μg poly(A$^+$) RNA from 24 DAP W64A (*Glb1-L/L*) *Glb1-0/0* embryos probed with cDNAs for *Glb1-S*, *Glb2*, and *L3*. B, Comparison of *Glb1* nuclear transcription in W64A (*Glb1-L/L*) and *Glb1-0/0* embryos. The data obtained for *Glb1* have been normalized to the level of *Glb2* transcripts in the two genotypes.

cates that the extremely low steady-state level of *Glb1* transcripts in *Glb1-0/0* embryos can not be attributed to a low amount of transcription of *Glb1-0*. Similar analysis of the *L3* gene indicates that the higher steady-state level of *L3* transcripts in *Glb1-0/0* embryos relative to *Glb1-L/L* embryos may be due to higher levels of transcription.

DISCUSSION

We present here characterization of nucleotide sequences corresponding to three allelic variants of the maize *Glb1* gene. The molecular basis for such polymorphism was determined by sequence comparisons of the three alleles. The *Glb1-L* allele differs from the *S* allele by a 36-bp imperfect duplication near the 3′ end of the protein coding sequence. This results in 12 additional amino acids at the corresponding position in the GLB1-L polypeptide relative to GLB1-S. There are three small deletions in the *Glb1-L* protein coding sequence relative to the *S* allele and another small deletion in *Glb1-S* relative to the *L* allele. The net result of these differences is that, relative to the *S* allele, the *L* allele encodes 9 additional amino acids resulting in a size difference of 1114 D in the primary translation products. This is relatively close to the experimentally determined difference of 2000 D observed by SDS-PAGE analysis of *in vitro* translation products (KRIZ and SCHWARTZ 1986). There are 20 single-base differences between the *Glb1-S* and *-L* alleles within the protein coding sequences. Most of these are silent changes as they result in only 7 amino acid replacements between the two proteins.

The defect in the *Glb1-0* null allele is due to an 11-bp insertion in the first exon, resulting in a translational frameshift which introduces a premature termination codon in the N-terminal region of the protein. Null alleles for several other genes encoding seed

proteins have been described, including translational frameshift mutants in the case of the soybean Kunitz trypsin inhibitor KTi3 allele (JOFUKU, SCHIPPER and GOLDBERG 1989), the *Phaseolus* phytohemagglutinin *Pdlec1* gene (VOELKER, STASWICK and CHRISPEELS 1986), and the soybean glycinin *gy4* "Raiden" allele (SCALLON, DICKINSON and NIELSEN 1987). The frameshift in each of these genes is due to a single-base mutation. In contrast, the translational frameshift in the *Glb1–0* allele is unusual in that it originates from a duplication of an adjacent sequence.

In addition to identifying the defect in *Glb1–0*, the present sequence analysis also allows for certain conclusions to be drawn concerning the origin of this allele. The *Glb1–0* allele contains the 36-bp insertion observed in the last exon in the *L* allele relative to the *S* allele. There are no instances of insertion/deletion differences between *Glb1-L* and *Glb1–0* within the protein coding regions other than that in the first exon which results in the translational frameshift. From the greater sequence similarity of, and the presence of the characteristic 36-bp insertion in, both the *Glb1-L* and *-0* alleles, the *0* allele must be more closely related to the *L* allele than to the *S* allele. The regions in *Glb1–0* corresponding to the protein coding sequences of *Glb1-L* show 16 single base changes relative to the *L* allele, as compared to the 20 single-base differences observed between *Glb1-L* and *Glb1-S*. In addition, the 5′ untranslated regions of the *Glb1-L* and *-0* alleles are identical, while that of *Glb1-S* contains an additional 11 nucleotides of coding sequence and three base substitutions relative to the *L* and *0* alleles. Given that *Glb1-L* and *Glb1–0* are more similar to each other than is either allele to *Glb1-S*, it is likely that *Glb1–0* is derived from *Glb1-L*. There are, however, seven single base positions within the protein coding sequences where the *S* and *0* alleles are the same and the *L* allele differs. These presumably represent changes which have occurred in the *L* allele subsequent to the divergence of the *L* and *0* alleles.

From primer extension analysis, the 5′ end of the transcript from each of the three alleles was determined to be at the adenine which is designated as position 1 in Figure 2. This is 57 bp upstream from the translation start site in the *S* allele and 46 bp upstream in the *L* and *0* alleles. As indicated above, the 11 additional nucleotides in the 5′ untranslated sequence of *Glb1-S* represents another polymorphism which distinguishes this allele from the *L* and *0* alleles. The length of the 5′ untranslated region in the three *Glb1* genes is similar to that observed for other plant genes (MESSING *et al.* 1983). The sequences surrounding the start site are similar to the higher plant consensus sequence of CTCATCA (JOSHI 1987).

The sequence TATAAAT at position −33 to −27 in all three alleles presumably represents a canonical TATA box (PROUDFOOT 1979). The relative position of this sequence with respect to the transcription start site is again similar to that observed for other plant genes (WALBOT and MESSING 1988). There is no obvious CAAT box (BREATHNACH and CHAMBON 1981) in the 5′ region of the gene but this is not unusual for plant genes (WALBOT and MESSING 1988). The sequence AAGGAGAG 134 bp upstream of the transcription start site resembles the AGGA box which may substitute for a CAAT box in the regulation of transcription (MESSING *et al.* 1983). The position in the *Glb1* genes at −134 is, however, further upstream from the transcription start site than the commonly found position of −80 to −100 (MESSING *et al.* 1983). Experiments are in progress to delimit actual regulatory regions involved in *Glb1* expression.

TABLE 1

Comparison of ABA-response elements in the wheat *Em* gene with maize *Glb1* sequences

Sequence designation[a]	Wheat *Em* gene[a]	Maize *Glb1* gene[b]
Em1a	−149 ACGTGGCGC	−118 ACGTGGCGAC
Em2	−125 CGAGCAG	−161 CGAGCCG
Em1b	−94 ACGTGCCGC	−76 ACGTAGCCGC

[a] From MARCOTTE, RUSSELL and QUATRANO (1989).
[b] Differences with the wheat sequences are underlined.

There are sequences in the 5′ region of the *Glb1* genes which are similar to the phytohormone abscisic acid (ABA) responsive elements in the wheat *Em* gene as described by MARCOTTE, RUSSELL and QUATRANO (1989). *Glb1* is regulated by ABA (KRIZ, WALLACE and PAIVA 1990), and it is likely that similar *cis*-acting elements are involved in the response of *Glb1* to ABA. A comparison of the *Em* sequences and the *Glb1* sequences is shown in Table 1. In all three cases, there are single base differences in the *Glb1* sequences as compared with the *Em* sequences. The relative positions of the three elements differs between the *Glb1* genes and the *Em* gene. Sequences similar to the Em2 element are present in the seed storage globulins wheat triticin and the α' subunit of soybean β-conglycinin (MARCOTTE, RUSSELL and QUATRANO 1989). Triticin is regulated by ABA (WILLIAMSON and QUATRANO, 1988) whereas the α' β-conglycinin subunit is not ABA-responsive (BRAY and BEACHY 1985). The Em2 element may thus be involved in regulation of gene expression in a seed specific manner (MARCOTTE, RUSSELL and QUATRANO 1989). The functional significance of these sequences in the maize *Glb1* gene remains to be determined.

A striking feature of the comparison of the three *Glb1* nucleotide sequences is the number of insertions within one allele relative to another. Many of these insertions are perfect or imperfect repeats of an adjacent sequence. This is the case for all five instances

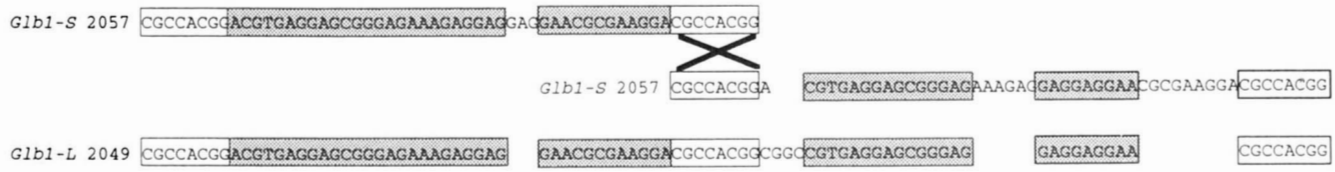


FIGURE 7.—Unequal crossover as a potential cause of the 36-bp imperfect duplication seen in the *Glb1-L* allele. In this example, the *Glb1-S* sequence has been mispaired with itself at the region of an 8-bp direct repeat (boxes). The predicted product of an unequal crossover at the mispairing is compared with the actual *Glb1-L* sequence. Regions of identity between the predicted product and the actual *Glb1-L* sequence are indicated by filled boxes.

of insertions in the protein coding regions of the *Glb1* alleles. Many of the insertions in the intron sequences are also perfect or imperfect repeats of adjacent sequences. Such insertions have been described in the introns of maize *Sh1* alleles (ZACK, FERL and HANNAH 1986) and in the single intron of maize *Bz1* alleles (FURTEK *et al.* 1988). Small duplications in one allele relative to another have also been reported in a random selection of maize RFLP clones (SHATTUCK-EIDENS *et al.* 1990).

As might be expected, the introns of the *Glb1* alleles differ to a greater extent than do the protein coding regions. Comparison of the *Glb1-L* and *Glb1-S* protein-coding sequences indicates the presence of 20 base substitutions and five instances of an insertion in one allele relative to the other. Based on the *Glb1-L* protein coding sequence of 1746 bp, these differences occur at frequencies of 1% and 0.2%, respectively. Within introns, however, both substitutions and insertions occur at a frequency of about 2%. The fact that such insertions are present within both the protein coding regions and the introns of the *Glb1* alleles may be due to a lesser degree of selection pressure against amino acid additions in genes encoding dispensable storage proteins as compared with genes encoding essential metabolic enzymes such as the *Sh1* gene product (ZACK, FERL and HANNAH 1986). The amino acid insertions which distinguish the *Glb1-L* and *-S* alleles result in changes in GLB1 protein structure with no apparent effects on seed development, maturation, or germination. In addition, the insertion in the *Glb1–0* allele, which results in a frameshift and subsequent premature termination of translation, has been maintained in the homozygous state with no loss of seed viability, again demonstrating the nonessential nature of the GLB1 protein. These observations indicate that the *Glb1* gene may serve as an excellent marker for analysis of genetic variation at the molecular level.

Short nucleotide insertions in a DNA sequence may arise in a variety of different ways. Unequal crossing over between homologous chromosomes during cell division has been shown to result in sequence duplications (ANDERSON and ROTH 1977). The 36 bp insertion in the last exon of *Glb1-L* to *Glb1-S* has features which suggest this rearrangement may have originated from an unequal crossover event. As stated above, this insertion is an imperfect duplication of the region just preceding it. An unequal crossover may have occurred at a mispairing of an eight bp direct repeat (CGCCACGG) which is found twice in the *Glb1-S* allele and three times in the *L* allele (Figure 7). Similarly, the eleven bp insertion found in the *Glb1–0* allele relative to the *S* and *L* alleles may have resulted from an unequal crossover at a mispairing of the sequence GAGG. Another possible origin of insertions which are perfect or imperfect duplications of adjacent sequences is transposable element "footprints." The excision of a plant transposable element leaves behind a duplication of the host sequence, with the size of the duplication being characteristic of the transposable element family (reviewed by DORING and STARLINGER 1986). SCHWARZ-SOMMER *et al.* (1985) proposed that transposable element footprints may play an important role in the generation of DNA sequence diversity.

Slipped-strand mispairing is another mechanism which has been proposed to result in the expansion of short repetitive units in a DNA sequence (LEVINSON and GUTMAN 1987). There are many short repeated sequences in the *Glb1* genes which may well have originated from slipped-strand mispairing. An example is the "gct" unit near the 5′ end of intron 3 (position 1049 in Figure 2). There are four of these units in this region in *Glb1-S*, three in the *L* allele, and five in the *0* allele. Similar small duplications found in one allele relative to another are also found in the coding regions of the *Glb1* alleles. An example is the GAG insertion found at position 343 in the first exon of *Glb1-S* (Fig. 2). MOORE (1983) has discussed slipped strand mispairing as a mechanism involved in generating length variation in introns. In the case of the *Glb1* alleles, it seems likely that this mechanism has generated short duplications in both the introns and exons.

It was previously determined that the steady-state level of *Glb1* transcripts was very low in embryos homozygous for the *Glb1–0* allele (BELANGER and KRIZ 1989). There are no significant differences, however, in the 5′ noncoding regions of the genes which might be expected to result in inefficient transcription of the *Glb1–0* gene (Figure 2). Nuclear run-on tran-

scription assays were therefore performed to determine if the transcription rate of the *Glb1–0* gene differed from that of the *Glb1-L* gene. The data from that experiment (Figure 6B) indicate that the low steady-state level of *Glb1* transcripts in *Glb1–0/0* embryos cannot be attributed to a low transcription rate. Although important regulatory elements likely exist further upstream than the 368 bp reported here, differences in these regions in the *Glb1–0* gene cannot be considered to be the major factor in the observed low steady-state level of *Glb1–0* mRNA. Because of the relatively high transcription rate of *Glb1–0,* the low steady-state level must therefore be due to instability of the *Glb1–0* messenger RNA. Similar results have been reported for other cases of frameshift mutations in plant genes which result in early termination of translation. Frameshift mutations in Kunitz trypsin inhibitor (JOFUKU, SCHIPPER and GOLDBERG 1989) and bean phytohemagglutinin (VOELKER, STASWICK and CHRISPEELS 1986) have been found to result in low steady-state levels of message but there were near normal transcription rates for these genes. As discussed by VOELKER, MORENO and CHRISPEELS (1990) the low steady state level of messages containing premature termination codons is likely due to enhanced cytoplasmic degradation of transcripts not protected by ribosomes.

We appreciate the excellent technical assistance provided by CHERYL GREEN. We thank A. H. C. HUANG for providing us with a cDNA clone for the lipid body protein L3, R. J. OKAGAKI for helpful discussions, and A. G. HEPBURN for assistance in preparing nucleotide sequence alignments. We appreciate the constructive criticism of this manuscript provided by LES DOMIER, PAUL SHAW and ROBERT RAMAGE. This research was supported in part by a grant from the U.S. Department of Agriculture Competitive Research Grants Office (No. 88–37262–3427) to A. L. K.

LITERATURE CITED

ANDERSON, R. P., and J. R. ROTH, 1977 Tandem genetic duplications in phage and bacteria. Annu. Rev. Microbiol. **31:** 473–505.

AVIV, H., and P. LEDER, 1972 Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69:** 1408–1412.

BEACH, L. P., D. SPENCER, P. J. RANDALL and T. J. V. HIGGINS, 1985 Transcriptional and post-transcriptional regulation of storage protein gene expression in sulfur-deficient pea seeds. Nucleic Acids Res. **13:** 999–1013.

BELANGER, F. C., and A. L. KRIZ, 1989 Molecular characterization of the major maize embryo globulin encoded by the *Glb1* gene. Plant Physiol. **91:** 636–643.

BRAY, E. A., and R. N. BEACHY, 1985 Regulation by ABA of β-conglycinin expression in cultured soybean cotyledons. Plant Physiol. **79:** 746–750.

BREATHNACH, R., and P. CHAMBRON, 1981 Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:** 349–383.

COX, R. A., 1968 The use of guanidinium chloride in the isolation of nucleic acids. Methods Enzymol. **12B:** 120–129.

DELLAPORTA, S. L., J. WOOD and J. B. HICKS, 1983 A plant DNA minipreparation: version II. Plant Mol. Biol. Reporter **1:** 19–21.

DORING, H.-P., and P. STARLINGER, 1986 Molecular genetics of transposable elements in plants. Annu. Rev. Genet. **20:** 175–200.

FURTEK, D., J. W. SCHIEFELBEIN, F. JOHNSTON and O. E. NELSON JR., 1988 Sequence comparisons of three wild-type *Bronze-1* alleles from *Zea mays*. Plant Mol. Biol. **11:** 473–481.

HUYNH, T. Y., R. A. YOUNG and R. W. DAVIS, 1985 Constructing and screening cDNA libraries in λgt10 and λgt11, pp. 245–300 in *DNA Cloning: A Practical Approach,* Vol. 1, edited by D. M. GLOVER. IRL Press, Oxford.

JOFUKU, K. D., R. D. SCHIPPER and R. B. GOLDBERG, 1989 A frameshift mutation prevents Kunitz trypsin inhibitor mRNA accumulation in soybean embryos. Plant Cell **1:** 427–435.

JOSHI, C. P., 1987 An inspection of the domain between putative TATA box and translation start site in 79 plant genes. Nucleic Acids Res. **16:** 6643–6653.

KINGSTON, R. E., 1989 Primer extension, Unit 4.8 in *Current Protocols in Molecular Biology,* edited by F. M. AUSUBEL, R. BRENT, R. E. KINGSTON, D. D. MOORE, J. G. SEIDMAN, J. A. SMITH and K. STRUHI. John Wiley & Sons, New York.

KODRZYCKI, R., R. S. BOSTON and B. A. LARKINS, 1989 The *opaque-2* mutation of maize differentially reduces zein gene transcription. Plant Cell **1:** 105–114.

KRIZ, A. L., 1989 Characterization of embryo globulins encoded by the maize *Glb* genes. Biochem. Genet. **27:** 239–251.

KRIZ, A. L., R. S. BOSTON and B. A. LARKINS, 1987 Structural and transcriptional analysis of DNA sequences flanking genes that encode 19 kilodalton zeins. Mol. Gen. Genet. **207:** 90–98.

KRIZ, A. L., and D. SCHWARTZ, 1986 Synthesis of globulins in maize embryos. Plant Physiol. **82:** 1069–1075.

KRIZ, A. L., M. S. WALLACE and R. PAIVA, 1990 Globulin gene expression in embryos of maize *viviparous* mutants. Evidence for regulation of the *Glb1* gene by ABA. Plant Physiol. **92:** 538–542.

LEVINSON, G., and G. A. GUTMAN, 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:** 203–221.

MARCOTTE JR., W. R., S. A. RUSSELL and R. S. QUATRANO, 1989 Abscisic acid-responsive sequences from the *Em* gene of wheat. Plant Cell **1:** 969–976.

MESSING, J., D. GERAGHTY, G. HEIDECKER, N-TH HU, J. KRIDL and I. RUBENSTEIN, 1983 Plant Gene Structure, pp. 211–228 in *Genetic Engineering of Plants,* edited by T. KOSUGE, C. P. MERIDITH and A. HOLLAENDER. Plenum Press, NY.

MOORE, G. P., 1983 Slipped-mispairing and the evolution of introns. Trends Biochem. Sci. **8:** 411–414.

OSTERMAN, J. C., 1988 An allele of the *Prot* locus in maize is a variant for the site of protein processing. Biochem. Genet. **26:** 463–474.

PROUDFOOT, N. J., 1979 Eukaryotic promoters? Nature **279:** 376.

PUCKETT, J. L., and A. L. KRIZ, 1991 Globulin gene expression in *opaque-2* and *floury-2* mutant maize embryos. Maydica **36:** 161–167.

SANGER, F., S. NICKLEN and A. R. COULSEN, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SCALLON R. J., C. D. DICKINSON and N. C. NIELSEN 1987 Characterization of a null-allele of the *Gy4* glycinin gene from soybean. Mol. Gen. Genet. **208:** 107–113.

SCHWARTZ, D., 1979 Analysis of the size alleles of the *Pro* gene in maize–evidence for a mutant protein processor. Mol. Gen. Genet. **174:** 233–240.

SCHWARZ-SOMMER, Z., A. GIERL, H. CUYPERS, P. A. PETERSON and H. SAEDLER 1985 Plant transposable elements generate the sequence diversity needed in evolution. EMBO J. **4:** 591–597.

SHATTUCK-EIDENS, D. M., R. N. BELL, S. L. NEUHAUSEN and T.

HELENTJARIS, 1990 DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. Genetics **126:** 207–217.

VANCE, V. B., and A. H. C. HUANG, 1987 The major protein from lipid bodies of maize. Characterization and structure based on cDNA cloning. J. Biol. Chem. **262:** 11275–11279.

VOELKER, T. A., J. MORENO and M. J. CHRISPEELS, 1990 Expression analysis of a pseudogene in transgenic tobacco: a frameshift mutation prevents mRNA accumulation. Plant Cell **2:** 255–261.

VOELKER, T., P. STASWICK and M. J. CHRISPEELS, 1986 Molecular analysis of two phytohemagglutinin genes and their expression in *Phaseolus vulgaris* cv. Pinto, a lectin-deficient cultivar of the bean. EMBO J. **5:** 3075–3082.

WALBOT, V., and J. MESSING, 1988 Molecular genetics of corn, pp. 389–429 in *Corn and Corn Improvement,* Ed. 3, edited by G. F. SPRAGUE and J. W. DUDLEY. American Society of Agronomy, Madison, Wisc.

WALLACE, N. H., and A. L. KRIZ, 1991 Nucleotide sequence of a cDNA clone corresponding to the maize *Globulin-2* gene. Plant Physiol. **95:** 973–975.

WILLIAMSON, J. D., and R. S. QUATRANO, 1988 ABA-regulation of two classes of embryo-specific sequences in mature wheat embryos. Plant Physiol. **86:** 208–215.

YANISCH-PERRON, C., J. VIEIRA and J. MESSING, 1985 Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:** 103–119.

ZACK, C. D., R. J. FERL and L. C. HANNAH, 1986 DNA sequence of a *Shrunken* allele of maize: evidence for visitation by insertional sequences. Maydica **31:** 5–16.

Communicating editor: W. F. SHERIDAN